BOVINE CASEINS: THREE DIMENSIONAL MOLECULAR MODELING

H.M. FARRELL, JR., .E.M. BROWN, and T.F. KUMOSINSKI
U.S. Department of Agriculture, Eastern Regional Research Center, Agricultural Research Service,
600 E. Mermaid Lane, Philadelphia, Pennsylvania, 19118 U.S.A.

BOVINE CASEINS: THREE DIMENSIONAL MOLECULAR MODELING

H.M. FARRELL, JR., .E.M. BROWN, and T.F. KUMOSINSKI
U.S. Department of Agriculture, Eastern Regional Research Center, Agricultural Research Service,
600 E. Mermaid Lane, Philadelphia, Pennsylvania, 19118 U.S.A.

ABSTRACT. Three dimensional (3-D) structures derived from X-ray crystallography are extremely important in elucidating structure-function relationships for many proteins. However, not all proteins can be crystallized. The caseins of bovine milk are one class of noncrystallizable proteins. The complete primary and partial secondary structures of these proteins are known, but homologous proteins of known crystallographic structures cannot be found. Therefore, sequence-based predictions of secondary structures were made and adjusted to conform with global secondary structures determined by Raman and FTIR spectroscopy. With this information 3-D structures for α_{s1}- and κ-caseins were constructed using the Sybyl-Mendyl molecular modeling programs. All unrefined structures showed good agreement with global biochemical and structural information concerning the caseins.

1. INTRODUCTION

At the heart of the skimmilk system is a unique biocolloid, the casein micelle. This colloidal complex is in dynamic equilibrium with its environment. Changes in the state of the casein micelle system occur during milk secretion and processing (9,20). Because of its innate importance to the milk system, the casein micelle has been studied extensively. From this large basis set of physical data, various investigators have attempted with varying degrees of success to formulate models which will adequately account for the properties of the casein micelle and allow for predictions of its behavior in real systems (9,20). These models often fall short because of a serious gap in the data basis set.

The combination of X-ray crystallography and molecular biology has contributed greatly to our understanding of the mechanisms of action of globular proteins. While the techniques of molecular biology are now being applied to casein, crystallographic structures will probably not be realized. It is this gap in our knowledge, the reconciliation of the physical chemical data for the caseins with molecular structure, that we have attempted to breech with the tools of three dimensional molecular modeling. These models are not an end in themselves; it is hoped that they will represent a new starting point in the examination and exploration of casein structure and function.

2. MATERIALS AND METHODS

2.1 Predictions of Secondary Structures

Selection of appropriate conformational states for the individual amino acid residues was accomplished by comparing the results of sequence-based predictive techniques, primarily those of Chou and Fasman (5), Garnier *et al.* (11) and Cohen *et al.* (6,7). Assignments of secondary structure (α-helix, ß-sheet or ß-turn) for the amino acid sequences were made when either predicted by more than one method, or strongly predicted by one and not predicted against by the others. In addition, because of the large number of proline residues in the caseins, proline-based turn predictions were made using the data of Benedetti *et al.* (2) and Ananthanarayanan *et al.* (1).

2.2 Molecular Modeling

Three dimensional structures of the individual caseins were approximated using molecular modeling methods with an Evans and Sutherland[1] PS390 interactive computer graphics display driven by Sybyl-Mendel (Tripos) software. Segments of the individual proteins ranging in length from 80 to 100 residues were constructed, and assigned ϕ and ψ angles characteristic of the respective predicted structures. All ω angles were assigned the conventional trans configuration. In addition, aperiodic structures are in the extended rather than totally random configuration. The Sybyl-Mendel subroutine "SCAN" was used, on the side chains only, to adjust torsional angles and relieve bad van der Waals contacts. The individual pieces were then joined together to produce the total polypeptide model.

3. RESULTS AND DISCUSSION

3.1 Rationale for Generation of Molecular Models from Secondary Structure Models

Various methodologies for sequence based secondary structural predictions are currently available (5,6,7,11). These methods have been applied to the caseins (8,14,18). In our initial approaches a sequence-based prediction was generated from each of three basis sets for α_{s1}- and κ-caseins. However, the relatively high proline contents of the caseins pose a problem. In attempting to generate a consensus sequence-based prediction, we first focused upon solving the proline problem in a way commiserate with known behavior of this residue in proteins and model peptides (1,2,19).

Recent Raman and FTIR spectroscopic analyses of the caseins have shown that up to 35% of the residues in caseins appear to be in ß- or γ-turns (3,4). To assess the probability that proline residues in the caseins might be located in reverse turns, we examined the tetrapeptides containing Pro in the second position for similarity with peptides known to form ß- and γ-turns (1,2). Most of the casein proline residues were predicted to be in a turn of some type (ß or γ). ß-Turns, other than those based on proline, were predicted by the Chou-Fasman (5) and Cohen et al. (7) methods. The total number of amino acid residues predicted to be in turns was only slightly in excess of the total predicted by spectroscopic methods (Table 1).

TABLE 1
COMPARISON OF ADJUSTED SEQUENCE BASED PREDICTIONS WITH SPECTROSCOPIC DATA

Sample		% Helix	% ß	% Turns	% Other
α_{s1}-Casein	Raman[3]	8 - 13	18 - 20	29 - 35	33 - 40
	Model	15	22	45	18
κ-Casein	FTIR[4]	10	43	15	32
	Model	16	27	37	20

In a similar fashion "consensus" scores for α-helix and ß-sheet were arrived at by choosing from those regions having the highest predicted probability of a given structure to yield values in accord with the FTIR and Raman data (3,4). In this case, all residues previously assigned to proline-based turns were eliminated first from consideration in α-helical or extended ß-structures. The net results of these calculations for κ- and α_{s1}-casein are compared with spectroscopic data in Table 1. Finally, all residues not included in these periodic structures were considered to be in an extended aperiodic conformation. The κ-casein model does not differ widely from that of Loucheux-Lefebvre and co-workers, (14) and with the exception of the ß- and γ-turns, the α_{s1}- model is comparable to that of Creamer et al. (8).

The secondary structural assignments which had been reconciled with FTIR and Raman spectroscopic data were used as a point of departure for the generation of three dimensional models. Idealized ϕ and ψ angles assigned initially for each structural element (1,2,19). Torsional angles were adjusted to remove unfavorable van der Waals contacts among the sidechains. The initial backbone conformation was maintained throughout this procedure.

3.2 Three Dimensional Molecular Model of κ-Casein

The computer model generated as described above for κ-casein is shown in Figure 1 with backbone only. In descriptive terms the protein can be thought of as being represented by a "horse and rider". The amino-terminal 110 to 120 residues represent the "horse" while the carboxy-terminal portion represents the "rider". Two distinct legs are seen in the "horse" portion of the model. These legs are generated by ß-sheet regions comprising residues 10 to 25; 29 to 34; 39 to 45 and 49 to 55, which are connected by γ- or ß-turns. The overall dimensions of the κ-casein monomer predicted by this model are 8.0 x 6.2 x 5.8 nm.

3.3 Chemistry of κ-Casein and the Three Dimensional Model

3.3.1 - Chymosin (Rennin)

κ-Casein differs from α_s-, and ß-caseins in that it is soluble over a broad range of calcium ion concentrations (23) and it would appear that κ-casein is the key to micelle structure in that it stabilizes the calcium insoluble α_{s1}- and ß-caseins. The action of chymosin (13) on the casein micelle is primarily the hydrolysis of the highly sensitive Phe-Met peptide bond (residues 105-106) of κ-casein (16). From studies of model peptides, it has been suggested that the residues lying between Pro-101 and Pro-109 occur in a ß-sheet structure (18). Using other predictive methods there is an equal chance that these residues could be in an α-helical conformation. In our model we show the α-helix which represents the "horse's bit" in our

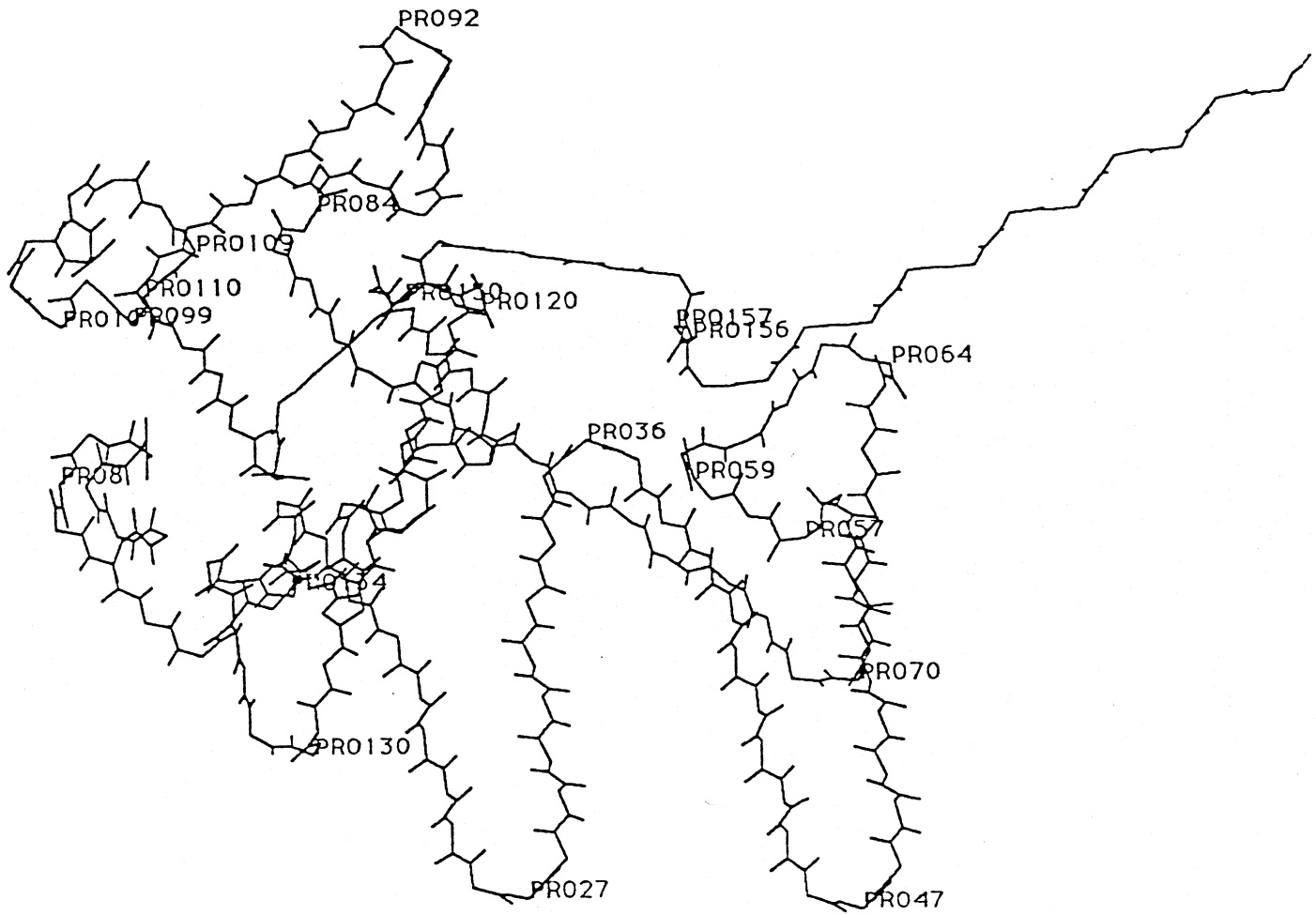

FIG. 1. Chain trace of κ-casein; prolines indicated.

descriptive "horse and rider" model. Whether the bond is in helical or sheet structure it is on the surface and readily accessible to chymosin

3.3.2 Hydrophobic Interactions

Earlier speculation (12) that κ-casein might be a linear amphiphile seems to be only partially true. The amino-terminal fifth of the molecule has a relatively high charge frequency (0.34), however, the net charge is zero and this part of the protein is also relatively hydrophobic. Residues 35 to 68 represent an exceptionally hydrophobic area with almost no charge. It is precisely within this region that the majority of the residues found in the "legged" structures of the κ-casein molecule occur. These non-stranded, highly hydrophobic ß-sheets make ideal sites for sheet-sheet interactions with other κ-casein molecules or with hydrophobic domains of α_s- and ß-caseins.

3.3.3 Sulfhydryl-Disulfide Interactions

κ-Casein contains two Cys residues. Whether these can form intra- or intermolecular disulfide bonds and the effects of such bonding on micelle stabilization have not been clearly established although studies on isolated κ-casein showed significant S-S bonding (14). It is interesting to note that Cys-11 is located between two segments predicted to be in α-helical conformations and is found on the "rider's" left in Figure 1, while Cys-88 is located in a predicted ß-turn on the opposite side. In our model both of these residues are located near the surface of the molecule and are directed away from each other. This could account for the ability of the κ-casein molecule (17,21) to form the inter-chain disulfide bonded polymers.

3.4 Three Dimensional Molecular Model of α_{s1}-Casein

The computer model, generated for α_{s1}-casein as described above is shown as backbone only in Figure 2 where it is displayed from carboxy- to amino-terminal (left to right). The best representation shows the molecule to be composed (right to left) of a hydrophilic amino-terminal portion, a segment of rather hydrophobic ß-sheet, the phosphopeptide segment and then a very hydrophobic carboxy-terminal domain.

3.5 Chemistry of α_{s1}-Casein and the Three Dimensional Model

3.5.1. Sites of Phosphorylation in α_{s1}-casein

α_{s1}-Casein B is a single polypeptide chain with 199 amino acid residues and a molecular weight of 23,619 (15). The α_{s1}-B molecule contains eight phosphate residues, all in the form of serine monophosphates. Seven of these phosphoserine residues are clustered in an acidic portion of the molecule bounded by residues 43 and 80 (the second fifth of the molecule from the amino-terminal end). This highly acidic segment contains 12 carboxylic acid groups as well as the seven phosphoserines. The model shows these residues to be located on ß-turns which is compatible with other known phosphorylated residues in proteins. This cluster forms a highly hydrophilic domain on the right shoulder of the molecule.

3.5.2. The α_{s1}-A Deletion

The rare α_{s1}-casein A genetic variant exhibits interactions which are highly temperature dependent. The A variant is the result of the sequential deletion of 13 amino acid residues bounded by residues 13 and 27; the majority of these deleted amino acids are apolar (10) but

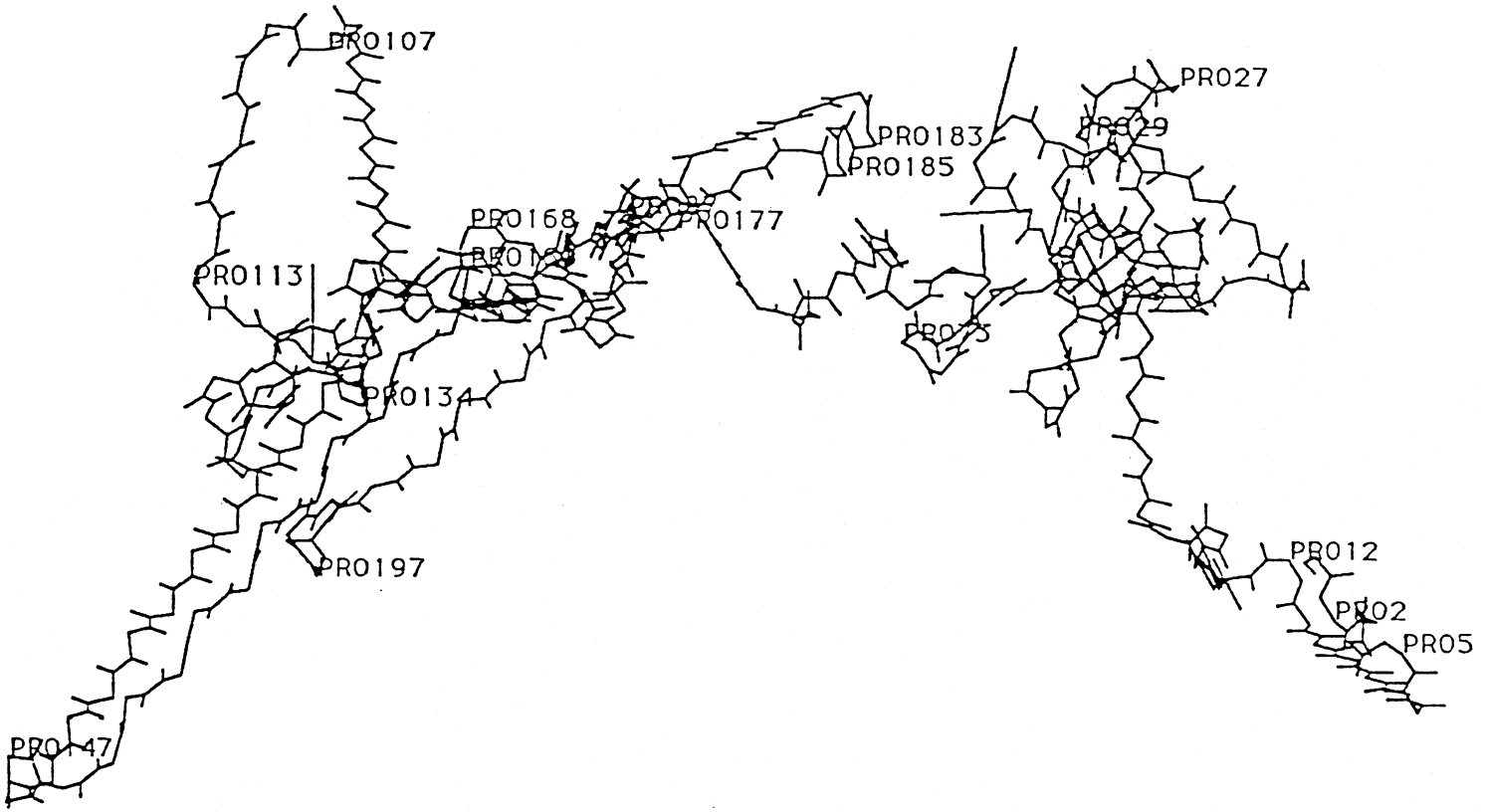

FIG. 2. Chain trace of α_{s1}-casein; prolines indicated.

Glu 18, Glu 14 and Arg 22 are also deleted. The deleted segment encompasses a region predicted to be in a ß-sheet. This sheet provides a spacer-arm between the hydrophilic amino-terminal region (five o'clock in Figure 2) and the phosphopeptide region. Deletion of this spacer-arm brings the hydrophilic section closer to the phosphate rich shoulder. In addition, the Phe-Phe bond (residues 23 and 24) is removed; this represents a major chymosin cleavage site and its absence may account for the poor quality products prepared from α_{s1}-casein A milks (22).

3.5.3. Hydrophobic Interactions

For α_{s1}-casein, the high degree of hydrophobicity exhibited by the segment containing residues 100 to 199 is probably responsible, in part, for the pronounced self-association of the α_{s1}-casein monomer in aqueous solution (20). This hydrophobic segment (residues 100 - 199) of α_{s1}-casein represents a hydrophobic domain and contains a segment of non-stranded ß-sheet reminiscent of those found in κ-casein.

3.6 Refinement of Casein Structures

It must be stressed that the structures presented in this work represent preliminary models. They have been partially refined to remove poor van der Waals side chain contacts, but they are not energy minimized structures. Thus electrostatic interactions, hydrogen bond formation, and backbone interactions are not taken into account. The refinement of these structures through the use of tools such as the Kollman Force Field will be the thrust of future work. However it is imperative to note that even this nonrefined molecular modeling technique (when combined with predicted secondary structures and spectroscopic results) can reveal important structure-function relationships. This information can in turn be used for designing new functional proteins using site directed mutagenesis.

REFERENCES

1. Ananthanarayanan, V.S., S.K. Brahmachari, and N. Pattabiraman. Arch. Biochem. Biophys. (1984) **232**, 482.
2. Benedetti, E., A. Bavoso, B.D. Blasio, V. Pavone, C. Pedone, C. Toniolo, and G.M. Bonora. Biopolymers (1983) **22**, 305.
3. Byler, D.M., H.M. Farrell, Jr. and H. Susi. J. Dairy Sci. (1988) **71**, 2622.
4. Byler, D.M. and H.M. Farrell, Jr. J. Dairy Sci. (1989) **72**, 1719.
5. Chou, P.Y. and G.D. Fasman. Adv. Enzymology (1978) **47**, 45.
6. Cohen, F.E., R.M. Abarbanel, I.D. Kuntz, and R.J. Fletterick. Biochemistry (1983) **22**, 4894.
7. Cohen, F.E., R.M. Abarbanel, I.D. Kuntz, and R.J. Fletterick. Biochemistry (1986) **25**, 266.
8. Creamer, L. K., T. Richardson, and D.A.D. Parry. Arch. Biochem. Biophys. (1981) **211**, 689.
9. Farrell, H.M., Jr. In Fundamentals of Dairy Chemistry, **3rd** Edition. N. Wong (ed.) (1988) Avi Press, Westport, CT p.461.
10. Farrell, H. M., Jr., T.F. Kumosinski, P. Pulaski, and M.P. Thompson. Arch. Biochem. Biophys. (1988) **265**, 146.
11. Garnier, J., D.J. Osguthorpe, and B.J. Robson. J. Mol. Biol. (1978) **120**, 97.
12. Hill, R.J. and R.G. Wake. Nature (1969) **221**, 635.